Revista de Contabilidade e Organizações
www.rco.usp.br

DOI: http://dx.doi.org/10.11606/issn.1982-6486.rco.2020.171481

Journal of Accounting and Organizations
www.rco.usp.br

# Inteligência artificial aplicada a pequenas empresas: o uso da engenharia automática de recursos e do aprendizado de máquina para um planejamento mais preciso

*Artificial intelligence applied to small businesses: the use of automatic feature engineering and machine learning for more accurate planning*

Alexandre Moreira Nascimento[a], Vinicius Veloso de Melo[b], Anna Carolina Muller Queiroz[a], Thomas Brashear-Alejandro[cd], Fernando de Souza Meirelles[d]

[a] *Universidade de Stanford*
[b] *The Wawanesa Mutual Insurance Company*
[c] *Universidade de Massachusetts*
[d] *Fundação Getúlio Vargas*

Palavras-chave
Inteligência artificial.
Engenharia automática de recursos.
Aprendizado de máquina.
Pequenas empresas.
Empresas locais.

## Resumo

O objetivo deste estudo é desenvolver um modelo preditivo que aumente a precisão do planejamento operacional de negócios usando dados de uma pequena empresa. A partir de técnicas de aprendizado de máquina (AM), são apresentadas estratégias de expansão, reamostragem e combinação que permitiram superar várias das limitações enfrentadas pelas pesquisas conduzidas até então. O estudo adotou uma nova técnica de engenharia de recursos que permitiu aumentar a precisão de um modelo preditivo, encontrando 10 novos recursos derivados dos originais, desenvolvidos automaticamente através das relações não-lineares encontradas entre eles. Por fim, foi criado um classificador com regras para prever, com alta precisão, a receita da pequena empresa. De acordo com os resultados apresentados, a abordagem proposta abre novas possibilidades para a pesquisa sobre a AM aplicada a pequenas e médias empresas.

*Keywords*
*Artificial intelligence.*
*Automatic feature engineering.*
*Machine learning.*
*Small business.*
*Local business.*

## *Abstract*

*The purpose of this study is to develop a predictive model that increases the accuracy of business operational planning using data from a small business. By using Machine Learning (ML) techniques feature expansion, resampling, and combination techniques, it was possible to address several existing limitations in the available research. Then, the use of the novel technique of feature engineering allowed us to increase the accuracy of the model by finding 10 new features derived from the original ones and constructed automatically through the nonlinear relationships found between them. Finally, we built a rule-based classifier to predict the store's revenue with high accuracy. The results show the proposed approach open new possibilities for ML research applied to small and medium businesses.*

Informações do artigo
Recebido: 27 de junho de 2020
Aprovado: 14 de setembro de 2020
Publicado: 14 de outubro de 2020

## Implicações Práticas

O estudo mostra como pequenas empresas, mesmo contanto com poucos dados, podem utilizar as técnicas disponíveis de aprendizado de máquina (AM) para prever diariamente sua receita total de vendas. Esse tipo de informação contribui para um planejamento mais preciso, melhorando o controle de estoque, reduzindo perdas de produto em caso de excesso, ou de receita em caso de falta.

Copyright © 2020 FEA-RP/USP. All rights reserved.

Autor correspondente: Tel. +55 (11) 3799-7755
E-mail: alexandremoreiranascimento@alum.mit.edu (A. M. Nascimento); vvdemelo@wawanesa.com (V. V. de Melo); acmq@stanford.edu (A. C. Muller Queiroz); Brashear@isenberg.umass.edu (T. Brashear-Alejandro); fernando.meirelles@fgv.br (F. de S. Meirelles)
Fundação Getulio Vargas, Escola de Administração de Empresas de São Paulo. Av. Nove de Julho, 2029 - Bela Vista, São Paulo/SP - 01313902, Brasil.

## 1 INTRODUÇÃO

A incerteza nas vendas é um grande desafio para as pequenas empresas (Lensink, Van Steen, & Sterken, 2005; Love & Hoey, 1990). Nessa direção, uma meta-análise dos fatores contextuais que impactam a relação entre o planejamento e o desempenho dessas organizações identificou que há falta de informação sobre o tempo levado para o fechamento de um negócio (Brinckmann, Grichnik, & Kapsa, 2010). Assim, pode-se concluir que o acesso a previsão de vendas se torna um fator preponderante para o sucesso de pequenos empreendimentos (Brinckmann et al., 2010).

Os recentes avanços na área de aprendizado de máquina (AM) – uma categoria de técnicas de inteligência artificial – podem contribuir para reduzir a incerteza nas vendas, prevendo-as a partir de dados disponíveis e ao encontrar relações complexas entre fatores diversos e transações anteriores. Por meio de AM, é possível modelar como esses elementos influenciam as vendas e fazer previsões diárias no caso de pequenas empresas, trazendo maior clareza para o planejamento e a operação do negócio, bem como aprimorando a definição de políticas, estratégias e táticas comerciais (Moore, 2008).

No entanto, poucos estudos têm se dedicado ao uso de técnicas avançadas de AM para fazer esse tipo de previsão em pequenas empresas, bem como em mercearias e supermercados. Uma busca por pesquisas em inglês no banco de dados Scopus usando a expressão booleana *[(“forecast”) AND (“supermarket” OR “grocery store”) AND (“daily”) AND (“demand” OR “revenue” OR “sales”)]* resultou em apenas 15 trabalhos, sendo que desses apenas 8 estão relacionados ao tema (Aburto & Weber, 2007; Berry, Helman, & West, 2020; Bousqaoui, Achchab, & Tikito, 2019; Deb, 2017; Kolassa, 2013; Slimani, El Farissi, & Achchab, 2017; Slimani, Farissi, & Al-Qualsadi, 2016; J. W. Taylor, 2011). Entretanto, nenhum desses estudos examina pequenas empresas ou adota engenharia automática de recursos.

Há uma lacuna na literatura sobre a aplicação de técnicas envolvendo AM em pequenos varejistas que atuam em lojas físicas. Como são técnicas que requerem muitos dados e variáveis para oferecer modelos úteis, acredita-se que sua aplicabilidade nesse tipo de negócio é restrita e, consequentemente, os proprietários seguem tomando a maioria das decisões confiando somente em sua intuição (Culkin & Smith, 2000; Fadahunsi, 2012).

Esse trabalho apresenta o uso de uma técnica de AM para a engenharia automática de recursos proposta recentemente – a Programação Kaizen (PK) – buscando superar a limitação causada pela escassez de dados. A pesquisa introduz uma abordagem criativa baseada no uso de dados secundários públicos para expandir o escasso conjunto de dados específicos de uma pequena mercearia, com objetivo de prever sua receita futura de vendas diárias e dar suporte ao gestor do negócio para aprimorar seu planejamento e tomada de decisão. Com isso, o estudo abre caminhos para futuros trabalhos na área.

Além dessa primeira seção introdutória, a próxima traz os antecedentes da transformação do setor de varejo e as oportunidades criadas a pequenos empreendimentos, seguida pela seção 3 que detalha a metodologia adotada no estudo. A seção 4 apresenta os resultados obtidos, discutidos na seção 5. Por fim, a seção 6 traz a conclusão do trabalho.

## 2 A TRANSFORMAÇÃO DO VAREJO – UMA OPORTUNIDADE PARA PEQUE-NAS EMPRESAS

O fenômeno da digitalização e os avanços no campo da Inteligência artificial estão causando a ruptura de muitas indústrias tradicionais (Stone et al., 2016; K. Taylor & Hanbury, 2018), entre elas, a do varejo (Gilbert, 2015). Empresas tradicionais com marcas fortes estão lutando para se adaptar e competir com empresas que nasceram com DNA digital (Loebbecke & Picot, 2015). Consequentemente, enquanto redes tradicionais como a Toys’R Us estão fechando suas portas (Isidore, Wattles, & Kavilanz, 2018) nos Estados Unidos, a Amazon vem expandindo sua presença por meio de lojas físicas (Soper, 2017).

Essa situação ameaça não só grandes empresas de varejo, que estão ficando para trás e falindo (K. Taylor & Hanbury, 2018) mesmo contando com consideráveis recursos econômicos e políticos. Ela afeta também os varejistas menores, e de forma mais aguda (Corkery, 2018). Na verdade, as pequenas lojas vêm sofrendo há anos com as dificuldades impostas pelos preços dos grandes fabricantes e poder de compra dos concorrentes de maior porte (Bowman, 2016). As grandes redes têm poder de barganha e conseguem operar com margens baixas, oferecendo preços mais competitivos. Os varejistas locais e de pequeno porte ficam em desvantagem, uma vez que operam com margens mais altas decorrentes de maiores despesas, estruturas de custos menos eficientes e menor volume de vendas (Draganska, Klapper, & Villas-Boas, 2010).

Para competir e sobreviver, alguns pequenos e médios varejistas têm se concentrado em nichos, como alimentos saudáveis e orgânicos produzidos com ingredientes de fornecedores locais (Dagevos, 2016). Eles optam por um foco específico, como o atendimento a uma clientela premium (Gil, Gracia, & Sanchez, 2000; Thompson, 1998; Van Doorn & Verhoef, 2011), ou privilegiando a qualidade ao preço (Doward, 2017), e/ou oferecendo o fator exclusividade como alavancagem, ou seja, adotam medidas que podem aumentar as margens, especialmente quando se trata de vendas e rotatividade em pequena escala. Por outro lado, à medida que os nichos se mostram viáveis, com crescimento da receita e expansão regional, grandes varejistas não raramente avançam sobre eles para conquistar outra fatia do mercado (Tu, 2016). Um exemplo nesse sentido é a Whole Foods nos EUA, uma loja de varejo que cresceu por meio de aquisições focadas em um nicho específico, sendo posteriormente comprada pela Amazon (Wingfield & de la Merced, 2017), que decidiu entrar no negócio de varejo físico e oferecer produtos saudáveis e com altas margens (Leswing, 2017). Portanto, as pequenas e médias empresas lutam para sustentar a vantagem competitiva mesmo em nichos de mercado, especialmente nos setores de varejo e supermercadista.

Historicamente, os grandes varejistas tradicionais têm acesso muito mais fácil às inovações de tecnologia da informação quando comparados aos pequenos e médios, considerando o porte dos investimentos necessários para adquirir e manter a infraestrutura computacional demandada. Essas condições favorecem as grandes empresas, que mantêm a vantagem competitiva mesmo sendo mais lentas para se adaptar às mudanças tecnológicas (Loebbecke & Picot, 2015). Já as pequenas e médias, mais rápidas e expostas a adaptações constantes (Li, Su, Liu, & Li, 2011) – fatores cruciais para seu crescimento e sobrevivência – têm acesso limitado a tecnologias inovadoras em virtude da escassez de recursos para investir.

No entanto, a revolução digital apoiada pela conectividade barata e acessível, a computação em nuvem, ferramentas de código aberto e os avanços da inteligência artificial e do AM, democratizou o acesso a tecnologias inovadoras poderosas (Dewhurst & Willmott, 2014; Müller & Bostrom, 2016). O impacto desse fenômeno – observado, por exemplo, em um número crescente de startups de tecnologia financeira (*fintech*) – é a redução das barreiras de acesso a ferramentas tecnológicas, favorecendo o planejamento, a tomada de decisão e a automação e otimização que antes estavam à mão apenas das grandes empresas (Dapp & Slomka, 2015). As pequenas, mesmo aquelas sem histórico de inovação, podem agora usar tecnologias para reduzir suas desvantagens, passando a se beneficiar competitivamente de suas características essenciais como a velocidade de adaptação e a facilidade de incorporar mudanças (Banks, 2013; Christensen & Bower, 1996; Cohen & Klepper, 1996; Davenport & Bibby, 1999; Jones, 2004). Entre as tecnologias disponíveis para que pequenos empreendimentos, nesse contexto de transformação, possam obter alguma vantagem competitiva, estão as técnicas baseadas em inteligência artificial (Burns, 2016; Gordon & Key, 1987; McFarlane, 1984), que permitem prever vendas e aprimorar a precisão do planejamento e otimização operacional.

A incerteza nas vendas é, na verdade, um dos maiores desafios das pequenas empresas (Lensink *et al.*, 2005) e um fator de risco no negócio de alimentos (Deb, 2017; Love & Hoey, 1990), podendo resultar em perdas diárias devido ao excesso ou falta de estoque (Deb, 2017; Ha, 1997). O excesso de estoque em lojas com produção própria de produtos naturais e saudáveis pode gerar perdas maiores, uma vez que a não utilização de conservantes e aditivos químicos nesses produtos reduz sua vida útil (Bonti-Ankomah & Yiridoe, 2006). Por outro lado, a falta de estoque pode levar à perda de receita pela falha em atender uma potencial demanda (Deb, 2017; Ha, 1997). Assim, esses negócios buscam estoques baixos, fornecedores com entregas frequentes e ajustes baseados na intuição e percepção dos atendentes, buscando evitar perda de produtos (Caspi, Pelletier, Harnack, Erickson, & Laska, 2016; Dunkley, Helling, & Sawicki, 2004). Essa estratégia subjetiva resulta em um incômodo percebido menor do que o incômodo da perda de vendas não previstas. Trata-se de uma política que reduz o risco, mas aumenta a sobrecarga operacional de tratamento de pedidos constantes (uma vez que os inúmeros ajustes finos não permitem uma programação prévia) (Andreyeva, Middleton, Long, Luedicke, & Schwartz, 2011). Ainda, o custo da adoção de entregas constantes não elimina perdas (nem a diminuição de receita por perda de vendas). Os pequenos empreendedores, então, operam em condições subótimas e tornam-se avessos a oferta de novos produtos (Andreyeva *et al.*, 2011).

Compreender os fatores que influenciam as vendas é fundamental no varejo, já que permite definir políticas, estratégias e táticas, bem como atuar nas variáveis que impactam esses fatores (Moore, 2008). Por exemplo, estimar o número médio de visitas à loja permite a otimização do planejamento da capacidade operacional, como número necessário de atendentes e caixas, otimização de estoque, planejamento de caixa e dívida, segurança e, no caso de empresas que fazem entregas, atende as questões de dimensionamento logístico (Lee & Whang, 2000; Moore, 2008). No entanto, as pequenas empresas ainda se baseiam principalmente na intuição dos proprietários para tomar a maioria de suas decisões (Culkin & Smith, 2000; Fadahunsi, 2012).

## 3 METODOLOGIA

Essa seção descreve a metodologia utilizada na pesquisa. A Figura 1 mostra um diagrama do protocolo utilizado, iniciando-se com a pergunta de pesquisa e a preparação dos dados, seguida de uma análise adotando a engenharia automática de recursos. O desequilíbrio da amostra de dados levou a realização de uma reamostragem com subsequente aplicação de um processo típico de indução de um classificador de AM, englobando o treinamento e teste do AM. Os resultados foram então relatados, apresentados na seção 4 desse artigo.

**Figura 1.** Processo de pesquisa
Fonte: elaborado pelos autores

### 3.1 Definição da pergunta de pesquisa

WA pergunta de pesquisa foi elaborada de forma a orientar a aplicação das técnicas para oferecer uma previsão capaz de auxiliar gestores de pequenas empresas a planejar seu negócio com maior precisão, e se resume na seguinte questão: *"Como prever a receita diária de vendas de uma pequena mercearia usando as poucas informações disponíveis?"*.

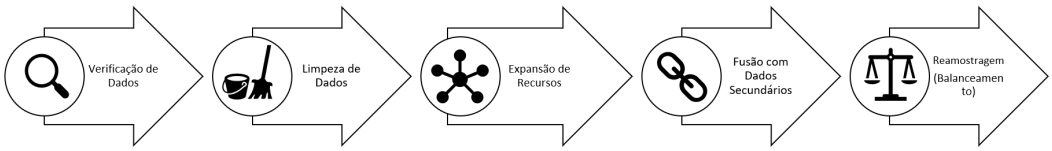


**Figura 2.** Preparação dos dados
Fonte: elaborado pelos autores

### 3.2 Preparação dos dados

Os dados coletados foram os valores da receita diária de vendas de uma loja ao longo de 1 ano, começando em 1 de agosto de 2016 e terminando em 31 de julho de 2017. A Figura 2 apresenta os vários procedimentos usados para prepará-los. O primeiro deles foi a verificação de dados, que consiste em sua formatação e na identificação de valores ausentes. Como os dados consistiam apenas na data e no faturamento do dia, foi necessário realizar um tratamento para obter novos recursos manualmente, a partir de atributos existentes. Esse tratamento tem importantes características e dá pistas sobre os modelos de AM a serem utilizados.

Um novo atributo – o dia da semana – foi gerado e adicionado, em formato numérico, ao conjunto de dados. Em seguida, a data foi particionada em mês e dia do mês, com o atributo original sendo substituído por dois novos atributos numéricos, obtendo-se úteis informações sazonais semanais e mensais. Por fim, para garantir que as informações sobre a sequência dos dias não fossem perdidas, foi adicionado um campo sequencial numérico para armazenar a sequência de cada um dos dias ao longo do ano.

Inicialmente, pretendia-se criar um modelo que pudesse prever com alta precisão a receita diária em um dia futuro, ou seja, o modelo deveria ser capaz de mostrar ao gerente da loja, por exemplo, que na próxima segunda-feira, o faturamento em vendas seria de R$7.325,00 e, quando comparada à receita real, a previsão apresentaria uma baixa taxa de erro. Porém, após muitas tentativas de utilização de técnicas de AM baseadas em regressão, percebeu-se que a escassez de dados não permitia atingir uma precisão adequada. Para superar essa limitação, optou-se por utilizar modelos de classificação de AM, o que ainda se mostrou útil, como validado com o gerente da empresa, uma vez que reduziu o intervalo das vendas diárias esperadas, proporcionando uma precisão maior para o planejamento em comparação com simples uso da intuição na gestão.

Assim, foi definido mais um novo atributo para substituir o faturamento diário – o nível de receita de vendas diárias –, de maneira que o modelo funcionasse com intervalos discretos de receita de vendas diárias como variável de destino. Este atributo foi obtido pela partição da faixa histórica da receita com vendas por dia (1,12e+03; 9,28e+03) em 5 níveis definidos por 5 faixas distintas de mesmo tamanho (1,63e+03): muito baixa {a = (1,12e+03; 2,75e+03]}; baixa {b = (2,75e+03; 4,38e+03]}; média {c = (4,38e+03; 6,01e+03]}; alta {d = (6,01e+03; 7,64e+03]}; e muito alta {e = (7,64e+03; 9,28e+03]}. Vale ressaltar que um pequeno número de níveis (3 por exemplo), aumentaria a precisão do modelo, mas não proporcionaria uma resolução adequada para o uso do gerente da loja. Um número ainda maior de níveis acaba por reduzir a precisão, proporcionando falsas expectativas ou certezas, o que prejudicaria o planejamento e poderia resultar em perdas indesejadas.

Posteriormente, dado o pequeno número de atributos em um modelo que explicasse a variável dependente (previsão do nível de receita de vendas diárias), os dados foram complementados com dados climáticos secundários. Por meio de um banco de dados histórico de informações climáticas (https://www.wunderground.com), foram obtidos os seguintes atributos: as informações diárias de temperatura (máxima e mínima), umidade relativa do ar e ocorrência de um evento climático (presença de vento, tempestade, chuva etc.). Com isso, os dados secundários foram combinados com os primários, obtendo-se um novo banco de dados. Por fim, outro recurso foi adicionado: uma flag indicando se o dia era feriado ou não. O Quadro 1 mostra os recursos (variáveis) incluídos na versão final do conjunto de dados usado como entrada para a etapa de engenharia automática de recursos. O modelo definido como alvo da análise é:

*NÍVEL-RECEITA-DIÁRIA = ƒ (SEQUÊNCIA, DIASEMANA, DIAMÊS, MÊS, FERIADO, TEMPMAX, TEMPMIN, UMIDADE, EVENTOCLIMA)*

**Tabela 1.** Variáveis (recursos) do conjunto de dados antes da engenharia automática de recursos

| Variável | Descrição |
|---|---|
| SEQUÊNCIA | Número único e sequenciado de 1 a 312 para informar a sequência dos dias da amostra |
| DIASEMANA | Número correspondente ao dia da semana (1 para domingo até 7 para sábado) |
| DIAMÊS | O dia do mês (1-31) |
| MÊS | O mês do ano (1-12) |
| FERIADO | 0 (zero) indica dia útil e 1 (um) indica feriado |
| TEMPMAX | Temperatura máxima do dia, em graus Celsius |
| TEMPMIN | Temperatura mínima do dia, em graus Celsius |
| UMIDADE | Percentual de umidade relativa |
| EVENTOCLIMA | 0 (zero) indica que não houve evento climático no dia (chuva, trovoada etc.); 1 (um) indica que houve evento climático no dia |

Fonte: elaborado pelos autores

Uma análise do conjunto de dados revelou um problema ainda maior relacionado a pequena quantidade de informações disponíveis: havia um desequilíbrio, com sub-representação de classes. O número de observações para cada classe foi de a=36, b=109, c=137, d=28 e e=2, o que demandou o uso de uma técnica de reamostragem para aumentar os exemplos das classes minoritárias a fim de equilibrar o conjunto de dados (Albisua *et al.*, 2013; Batuwita & Palade, 2010; Estabrooks, Jo, & Japkowicz, 2004; Ramentol *et al.*, 2012). Essa estratégia foi usada apenas para forçar o algoritmo de AM a incluir essas classes no modelo.

### 3.3 Engenharia automática de recursos

Técnicas de classificação, como o uso de algoritmos para árvores de decisão, por exemplo, buscam regras com base em uma gama de atributos para dividir o espaço multidimensional em diferentes regiões ou classes. No caso dos algoritmos de árvore de decisão, isso é feito por meio da divisão de cada nó interno de acordo com algum critério, como a divisão que dá a entropia mínima em termos de distribuição de classes (quanto menor a entropia, mais separadas as classes). Uma sequência de divisões pode ser convertida em uma regra de decisão. A Figura 3 ilustra uma árvore de decisão hipotética com o conjunto de regras e o melhor ajuste na divisão do plano XY com base em x e y, para separar as duas classes diferentes (x e círculo pintado de vermelho).

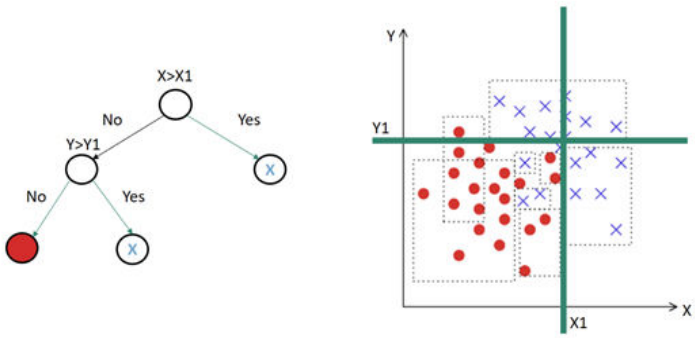


**Figura 3.** Classificação dos dados na árvore de decisão
Fonte: elaborado pelos autores

Uma característica clara dessa abordagem é que os atributos são avaliados de forma independente, resultando em divisões ao longo dos eixos e ignorando possíveis relações entre as variáveis. Para lidar com esse problema, pode-se executar a engenharia de recursos manual ou automática, combinando esses atributos. A técnica de engenharia automática de recursos utilizada nesta pesquisa é denominada Programação Kaizen (PK) (De Melo, 2014).

A PK é uma técnica iterativa baseada em uma abstração computacional da metodologia Kaizen usando o ciclo Planejar-Desenvolver/executar-Checar/verificar-Agir (PDCA). A PK divide um problema em partes para que possa ser examinado, sendo que cada uma é avaliada separadamente (obtendo-se uma solução parcial), para então ser melhorada e reunida em uma solução completa. Assim, a qualidade das contribuições parciais é avaliada junto com a da solução completa, verificando-se quais das parciais foram importantes e estabelecendo um ranking. As partes da solução que não geram contribuições relevantes podem ser descartadas, enquanto as demais são mantidas no processo de melhoria.

A PK é uma abordagem híbrida que combina um algoritmo de busca global (como um Algoritmo Evolucionário), e uma técnica de busca local eficiente para construir o modelo. Nessa programação, as técnicas global e local funcionam de maneira diferente das abordagens tradicionais, ou seja, o principal algoritmo na PK é a busca local (que constrói a solução completa), enquanto a global tenta apenas identificar regiões promissoras (as soluções parciais) do espaço de busca.

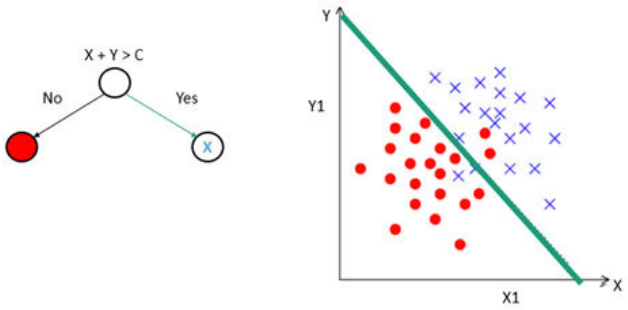


**Figura 4.** Regras de classificação dos dados baseadas na combinação dos recursos
Fonte: elaborado pelos autores

Em síntese, o algoritmo de busca global procura relacionamentos úteis entre os atributos do conjunto de dados, enquanto o local constrói o modelo completo usando os novos relacionamentos. A busca global usada aqui é uma técnica de computação evolutiva chamada Programação Genética (PG) que desenvolve soluções (indivíduos) cujos genes são expressões matemáticas. Por exemplo, suponha que o conjunto de dados tenha os atributos X e Y e as expressões disponíveis sejam +, -, * e /. Nesse caso, um possível indivíduo (ind) gerado por PG pode ser descrito como ind = X + Y. A PG tem uma população de indivíduos que evoluem por muitas gerações, otimizando uma determinada função objetiva, ou seja, maximizando a importância de cada novo atributo em um classificador. A técnica de classificação, no nosso caso, é uma árvore de decisão que fornece a importância de cada atributo utilizado para construir o modelo. Quanto maior a importância, melhor é o atributo e, consequentemente, melhor será o indivíduo na população. Dessa forma, a PG pode desenvolver melhores recursos que resultarão em melhores árvores de decisão.

Em outras palavras, a PK encontra fórmulas combinando os atributos que podem dividir o espaço com sabedoria e trazer melhores resultados de classificação. Ao encontrar essas fórmulas, há um ganho de informação (Ganho de Entropia = Entropia Antes – Entropia Depois), que pode fornecer recursos adicionais (expansão de recursos) ou pode substituir um conjunto de diferentes recursos por apenas um que esteja baseado em uma fórmula que combina todos os outros (redução de recursos). A Figura 4 mostra um exemplo em que a árvore de decisão usa o novo atributo descoberto pela PG e encontra o melhor valor de divisão (C) para separar as classes. Pode-se observar que a solução não só é melhor do que na Figura 3, mas é também menor.

### 3.4 Construindo novos recursos

Existem duas etapas na PK. Primeiro, a programação deve descobrir recursos que são de fato significativos para o classificador e, em seguida, cria-se um novo conjunto de dados anexando os novos recursos aos originais. Nessa pesquisa, destaca-se o fato de que o objetivo final é prever o nível da receita diária de vendas por meio de uma classificação sendo que a variável dependente original é um número real. Assim, foi considerada uma estratégia em que a PK desenvolve novos recursos em uma regressão usando um algoritmo de mínimos quadrados ordinários. Posteriormente, um algoritmo de classificação usa esses novos recursos com uma variável dependente discreta, resultando na classificação desejada.

Na PK foram realizadas 50 execuções independentes para desenvolver 10 novos recursos. As funções disponíveis para criar um novo recurso eram +, -, *, /, log, sqrt, abs (módulo) e 1/x, onde x pode ser um valor ou uma expressão. Na etapa do PDCA de checar/verificar, a PK gerou cinco novos recursos de cada um dos 10 originais (padrão). Em seguida, a programação remove os recursos com uma co-correlação de > 0,9 (mantendo no conjunto de dados aquele que apareceu primeiro) e usa o algoritmo de mínimos quadrados ordinários para criar um modelo e calcular a importância de cada recurso não correlacionado. Finalmente, os 10 melhores recursos são escolhidos para o novo modelo de teste. Se a qualidade, medida pelo erro quadrático médio, do modelo que usa os novos recursos for melhor do que a do modelo padrão, esse último é atualizado. Este ciclo é repetido 2.000 vezes para cada execução e, ao final do processo, selecionamos o conjunto de características de cada execução que apresentou a melhor qualidade.

### 3.5 Técnica de classificação

A classificação empregou o algoritmo PART na ferramenta Weka ML versão 3.6.15. O PART é um classificador que cria uma lista de regras de decisão empregando uma abordagem de divisão e conquista, construindo uma árvore de decisão parcial em cada iteração e transformando a "melhor" folha em uma regra. Escolhemos esse tipo de algoritmo porque suas regras geralmente resultam em caixas-brancas, enquanto algoritmos mais sofisticados, como redes neurais artificiais, resultam em caixas-pretas. Mesmo que as redes neurais artificiais sejam geralmente capazes de alcançar uma melhor qualidade de previsão, o fato de se apresentarem como caixas-pretas faz com que não sejam capazes de oferecer insights compreensíveis para humanos sobre a relação entre a variável dependente e as independentes. Por outro lado, o PART cria regras com base nas características descobertas pela PK e, embora elas resultem em caixas-brancas, alguns dos novos recursos podem não ser interpretáveis, tornando o resultado final uma caixa-cinza. Mesmo assim, existe uma relação útil entre os recursos originais, quando pertencem ao único novo recurso.

### 3.6 Treinando o classificador

Dado que o conjunto de dados de classificação não estava balanceado, empregamos o filtro de reamostragem supervisionada em Weka para ajustar o conjunto de dados e forçar a geração de regras capazes de abranger todas as classes. Caso contrário, as classes pequenas seriam "absorvidas" pelas maiores, já que a técnica prefere um modelo mais simples para melhor generalização ao invés de criar regras para classificar corretamente apenas algumas observações (especialização). A configuração, escolhida após análise empírica da distribuição, foi de 0,95 de viés e 250% do tamanho da amostra e nenhum outro filtro foi usado. Quanto à técnica de classificação, utilizamos o PART com sua configuração padrão na ferramenta.

### 3.7 Testando o classificador

Em nossa abordagem, o conjunto de dados aumentado foi usado no treinamento, enquanto o conjunto de dados original (não balanceado) foi usado para testar o classificador resultante. Esse procedimento é diferente de usar o conjunto do treinamento para a testagem, o que potencialmente forneceria uma previsão perfeita (Farhadi, 2018).

## 4 RESULTADOS

Os melhores recursos descobertos pela PK, ordenados dos mais para os menos importantes, são mostrados abaixo. Fica evidente que existem estruturas repetidas em diferentes recursos, como 1/log (log (<F>)) e log (<F>/ log (<F>)), onde <F> se refere a alguns dos originais. Embora semelhantes, os recursos possuem distribuições bastante diferentes (Figura 5).

O resultado do conjunto de testes utilizando o PART é mostrado na Tabela 2 e a matriz de confusão é apresentada na Tabela 3. Pode-se observar que, ao contrário do que foi demonstrado nos testes preliminares, os resultados apontam para uma boa diferenciação de classes e um grau marcante de sucesso.

**Tabela 2.** Resultados do PART usando a configuração padrão

| | |
|---|---|
| Amostras corretamente classificadas | 253 (81,09%) |
| Amostras incorretamente classificadas | 59 (18,91%) |
| Estatística Kappa | 0,7249 |
| Erro médio absoluto | 0,0803 |
| Raiz do erro quadrático médio | 0,2687 |
| Erro relativo absoluto | 25,09% |
| Raiz do erro quadrático relativo | 67,15% |
| Número total na amostra | 312 |

Fonte: elaborado pelos autores

**Tabela 3.** Matriz de confusão

| a | b | c | d | e | ← classificado como |
|---|---|---|---|---|---|
| 36 | 0 | 0 | 0 | 0 | a = (1,12e+03;2,75e+03] |
| 3 | 89 | 12 | 4 | 1 | b = (2,75e+03;4,38e+03] |
| 1 | 28 | 98 | 9 | 1 | c = (4,38e+03;6,01e+03] |
| 0 | 0 | 0 | 28 | 0 | d = (6,01e+03;7,64e+03] |
| 0 | 0 | 0 | 0 | 2 | e = (7,64e+03;9,28e+03] |

Fonte: elaborado pelos autores

Nota: Preto – Classificação correta; Cinza – Classificação incorreta)

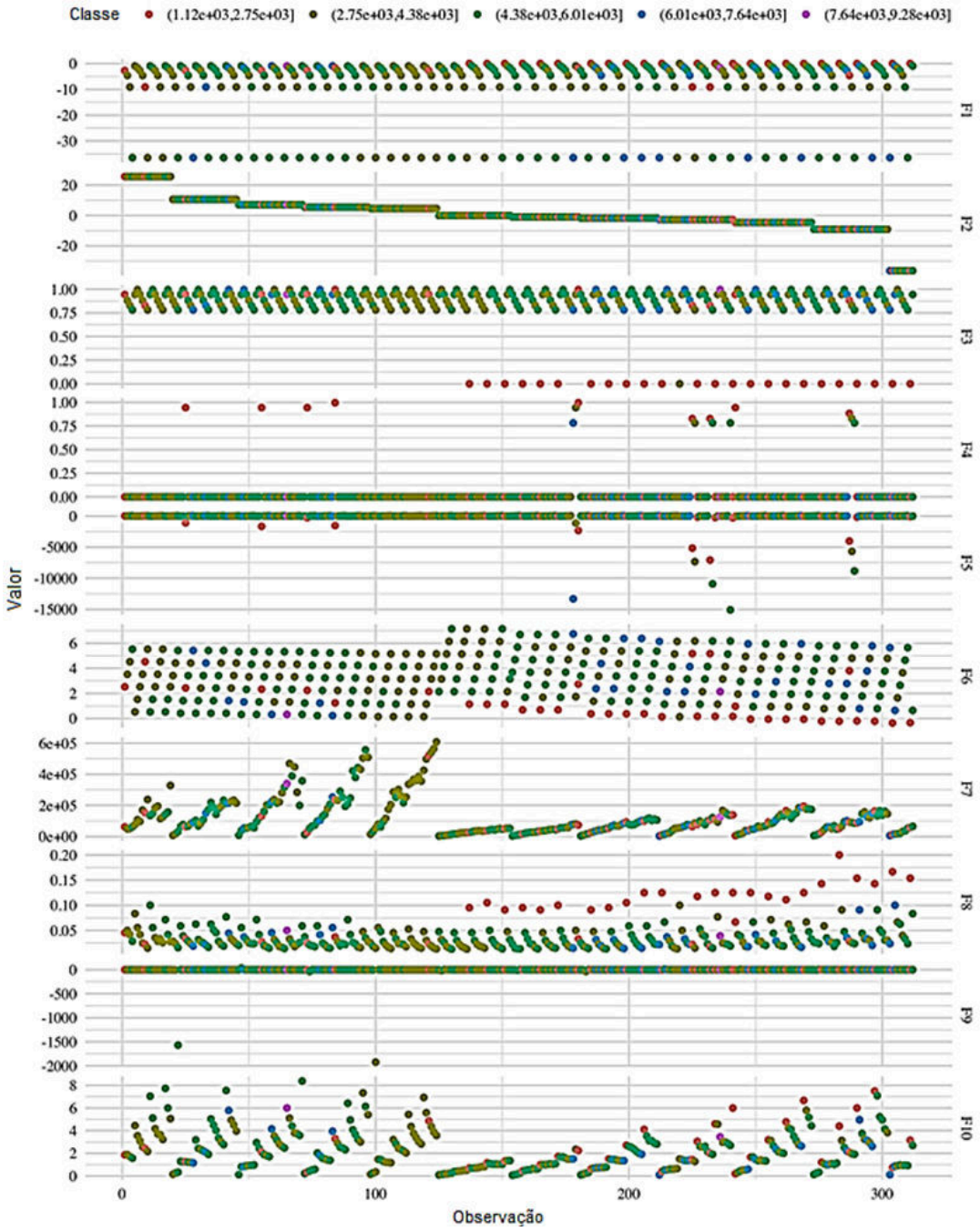


**Figura 5.** Plotagem de cada novo recurso para as 312 observações no conjunto de dados
Fonte: elaborado pelos autores
Nota: verificar o Apêndice para mais informações.

A Tabela 4 mostra a precisão da classificação realizada para cada classe (nível de receita diária de vendas). As classes 'a', 'd', 'e' (muito baixa, alta e muito alta) tiveram 100% das classificações feitas corretamente. As classes 'b' e 'c' (baixa e média) tiveram 82% e 72%, respectivamente, das classificações feitas corretamente. De 312 dias, 59 (18,9%) foram classificados incorretamente.

**Tabela 4.** Precisão da classificação por classes

| Classes | Precisão |
|---|---|
| a = (1,12e+03;2,75e+03] | 100% |
| b = (2,75e+03;4,38e+03] | 82% |
| c = (4,38e+03;6,01e+03] | 72% |
| d = (6,01e+03;7,64e+03] | 100% |
| e = (7,64e+03;9,28e+03] | 100% |

Fonte: elaborado pelos autores

## 5 DISCUSSÃO

A aplicação do método usado neste estudo demonstrou que o nível de receita de vendas diárias pode ser corretamente previsto em 81% dos casos durante a execução do conjunto de dados de teste. Portanto, o modelo baseado nos recursos (fórmulas) descobertos pela PK previu corretamente os níveis de receita em 253 de 312 dias. Ao saber o nível de receita de vendas diário com antecedência, o gestor pode estimar o número de clientes em um determinado dia, simplesmente dividindo o limite inferior e o limite superior da faixa de receita no nível (ou a média de ambos) conforme previsto pelo ticket médio (em torno de R$ 50,00 por transação).

O número de clientes que visitam a loja por dia pode ser usado para planejar a produção própria e o estoque necessário quando da comercialização de produtos de terceiros. O gerente pode calcular o número médio de cada produto vendido por cliente por dia em um período de dias anteriores, e usar essas relações para calcular as vendas esperadas nos dias seguintes.

O modelo proposto conseguiu explicar com maior propriedade os extremos do que os níveis intermediários de faturamento diário, mas também foi capaz de atingir um nível de precisão razoável (>70%) para esses últimos. Como os extremos são os níveis mais baixos e mais altos de receita diária de vendas, o modelo pode ajudar a reduzir grandes perdas de receita e custos de manutenção, prevendo melhor as necessidades de estoque diárias e semanais, bem como orientar a programação de produção e a escala de funcionários. Ele também fornece uma ferramenta poderosa para aumentar as habilidades dos gerentes de pequenas empresas em reconhecer e mitigar situações extremas.

Um aspecto importante do nosso estudo está relacionado à complexidade dos recursos descobertos pela PK. Há uma compensação na técnica entre a compreensibilidade das fórmulas descobertas e a precisão da previsão, ou seja, para maximizar a precisão das classificações, a PK foi configurada para adotar um alto nível de complexidade nas fórmulas geradas. A vantagem é o bom resultado na classificação, enquanto a desvantagem é que as fórmulas são difíceis de entender e de serem utilizadas racionalmente por gestores de pequenos empreendimentos. Por outro lado, pode-se decidir usar apenas operações aritméticas e limitar o tamanho das expressões numéricas, obtendo características mais simples.

Mesmo com sua alta complexidade, porém, essas fórmulas podem servir como ferramentas úteis para o gerente, podendo ser programadas numa planilha ou simples programa de computador. Em seguida, o gestor simplesmente fornece os valores de entrada (data, informações de previsão do tempo e se é feriado) e obtém o cálculo automaticamente. Dessa maneira, o algoritmo de classificação pode prever o nível de receita diária de vendas.

Os gestores podem também usar a ferramenta como um simulador para entender o impacto de diferentes fatores nos níveis diários de receita de vendas. As informações previstas podem ser usadas para apoiar o ajuste fino da administração de vários aspectos operacionais, como a produção própria, a quantidade e frequência de pedidos, as escalas de funcionários e outras alocações de recursos. Por exemplo, os funcionários podem dedicar-se a tarefas relacionadas a organização do espaço em dias com poucos compradores, e focar no atendimento ao cliente em dias de alta demanda. Além disso, os gestores podem influenciar o número de visitas à loja nos dias em que se espera que as vendas diárias sejam baixas. Por exemplo, ao antecipar que as baixas temperaturas reduzirão as vendas diárias, a loja pode enviar e-mails aos clientes dias antes, informando que oferecerá chocolate quente grátis em dias onde a previsão indicar menor movimento, ou e-mails em dias muito quentes e com altos níveis de umidade relativa, oferecendo sorvete ou suco orgânico gratuitamente.

## 6 CONCLUSÃO

Esse estudo propõe um novo método de previsão de faturamento diário que utiliza um conjunto de dados exclusivo e é aplicado a uma pequena empresa, demonstrando como as sofisticadas técnicas de AM disponíveis podem ser usadas por pequenas empresas para melhorar seu desempenho. As ferramentas permitem que os gestores ampliem a precisão do planejamento, minimizando perdas de estoque e reduzindo as perdas de receita por falta de estoque quando há picos de demanda. Os resultados obtidos mostram como uma pequena empresa, mesmo com dados limitados e escassos, pode usar técnicas de inteligência artificial para prever a receita total de vendas dia-a-dia, gerando informação que pode ser usada para aferir o número de clientes. O impacto da escassez e limitação de dados mencionado no presente estudo pode ser reduzido combinando conjuntos de dados externos (dados secundários, como o clima, por exemplo) e realizando expansão de recursos, engenharia automática de recursos e reamostragem para equilibrar os dados. Conforme observado ao longo do artigo, as informações previstas podem ser usadas para apoiar a gestão no ajuste fino de muitos aspectos operacionais, incluindo a própria produção, antecipando o tamanho dos pedidos e a frequência dos fornecedores, bem como a escala de funcionário e a agenda de entregas. Embora o conjunto de dados usado seja de uma pequena mercearia local focada em alimentos orgânicos e saudáveis, o método proposto pode ser usado para restaurantes e outros negócios de pequeno porte.

Os resultados indicam que a combinação inovadora dos métodos e técnicas aqui empregada é promissora. O teste realizou uma previsão correta dos níveis de receita de vendas por dia em 253 dias (81%) do período total de 312 dias analisados. Pode-se observar que as sofisticadas técnicas de AM estão sim ao alcance de pequenas empresas, que podem se beneficiar do valor adicionado pela inteligência artificial mesmo considerando a escassez e limitações de dados nesse tipo de organização. A PK teve um papel importante ao descobrir novos recursos capazes de aprimorar a precisão da classificação, demonstrando que essa programação pode representar uma poderosa ferramenta para a análise de dados de negócios.

O presente estudo apresenta limitações que abordamos a partir de nossas competências multidisciplinares, adotando a seleção de técnicas apropriadas de ciência de dados. Dentre essas limitações estão o pequeno conjunto de dados (apenas 312 amostras), a baixa representabilidade de algumas faixas de faturamento diárias (causando desequilíbrio na base de dados) e, por fim, os poucos atributos incluídos na base original. Assim, recomenda-se a realização de estudos sem essas limitações, no sentido de aprofundar os resultados aqui relatados.

Finalmente, o uso apropriado de técnicas de inteligência artificial por pequenas empresas representa um campo fértil para pesquisas com potencial impacto social, permitindo ampliar a competitividade dessas organizações. Ressalta-se ainda a recomendação por mais pesquisas que possam tanto superar as limitações deste estudo, usando outras fontes de dados secundários ou coletando mais dados primários por meio de técnicas simples que incentivam os clientes a fornecer mais informações à empresa. Por fim, recomenda-se que pesquisas semelhantes sejam realizadas com pequenos empreendimentos de diferentes setores para aumentar o repertório de dados disponíveis sobre o assunto.

## REFERÊNCIAS

Aburto, L., & Weber, R. (2007). A sequential hybrid forecasting system for demand prediction. *Lecture Notes in Computer Science*, 4571, 518–532. DOI: https://doi.org/10.1007/978-3-540-73499-4_39

Albisua, I., Arbelaitz, O., Gurrutxaga, I., Lasarguren, A., Muguerza, J., & Pérez, J. M. (2013). The quest for the optimal class distribution: an approach for enhancing the effectiveness of learning via resampling methods for imbalanced data sets. *Progress in Artificial Intelligence*, 2(1), 45–63. DOI: https://doi.org/10.1007/s13748-012-0034-6

Andreyeva, T., Middleton, A. E., Long, M. W., Luedicke, J., & Schwartz, M. B. (2011). Food retailer practices, attitudes and beliefs about the supply of healthy foods. *Public Health Nutrition*, 14(6), 1024–1031. DOI: https://doi.org/10.1017/S1368980011000061

Banks, G. P. (2013). *Exploring small-business change and strategic adaptation in an evolving economic paradigm*. Dissertação de doutorado. Walden University.

Batuwita, R., & Palade, V. (2010). Efficient resampling methods for training support vector machines with imbalanced datasets. In: *The 2010 International Joint Conference on Neural Networks* (IJCNN) (pp. 1–8).

Berry, L. R., Helman, P., & West, M. (2020). Probabilistic forecasting of heterogeneous consumer transaction–sales time series. *International Journal of Forecasting*, 36(2), 552–569. DOI: https://doi.org/10.1016/j.ijforecast.2019.07.007

Bonti-Ankomah, S., & Yiridoe, E. K. (2006). Organic and conventional food: a literature review of the economics of consumer perceptions and preferences. *Organic Agriculture Centre of Canada*, 59, 1–40.

Bousqaoui, H., Achchab, S., & Tikito, K. (2019). Machine learning applications in supply chains: Long short-term memory for demand forecasting. *Lecture Notes in Networks and Systems*, 49, 301–317. DOI: https://doi.org/10.1007/978-3-319-97719-5_19

Bowman, J. (2016). *Walmart's neighborhood market is crushing the competition*. Business Insider. Retrieved May 15, 2018, from: http://www.businessinsider.com/walmarts-neighborhood-market-is-crushing-the-competition-2016-8

Brinckmann, J., Grichnik, D., & Kapsa, D. (2010). Should entrepreneurs plan or just storm the castle? A meta-analysis on contextual factors impacting the business planning--performance relationship in small firms. *Journal of Business Venturing*, 25(1), 24–40. DOI: https://doi.org/10.1016/j.jbusvent.2008.10.007

Burns, P. (2016). *Entrepreneurship and small business*. Palgrave Macmillan Limited.

Caspi, C. E., Pelletier, J. E., Harnack, L., Erickson, D. J., & Laska, M. N. (2016). Differences in healthy food supply and stocking practices between small grocery stores, gas-marts, pharmacies and dollar stores. *Public Health Nutrition*, 19(3), 540–547. DOI: https://doi.org/10.1017/S1368980015002724

Christensen, C. M., & Bower, J. L. (1996). Customer power, strategic investment, and the failure of leading firms. *Strategic Management Journal*, 197–218.

Cohen, W. M., & Klepper, S. (1996). Firm size and the nature of innovation within industries: the case of process and product R&D. *The Review of Economics and Statistics*, 232–243.

Corkery, M. (2018). *Grocery Wars Turn Small Chains Into Battlefield Casualties*. The New York Times. Retrieved May 15, 2018, from: https://www.nytimes.com/2018/03/26/business/grocery-wars-small-chains.html

Culkin, N., & Smith, D. (2000). An emotional business: a guide to understanding the motivations of small business decision takers. *Qualitative Market Research: An International Journal*, 3(3), 145–157. DOI: https://doi.org/10.1108/13522750010333898

Dagevos, H. (2016). *Beyond the Marketing Mix: Modern Food Marketing and the Future of Organic Food Consumption*. In: The Crisis of Food Brands: Sustaining Safe, Innovative and Competitive Food Supply, 255.

Dapp, T., & Slomka, L. (2015). Fintech reloaded - Traditional banks as digital ecosystems. *Publication of the German Original*, disponível em: https://www.deutschebank.nl/nl/docs/Fintech_reloaded_Traditional_banks_as_digital_ecosystems.pdf

Davenport, S., & Bibby, D. (1999). Rethinking a national innovation system: The small country as' SME'. *Technology Analysis & Strategic Management*, 11(3), 431–462.

De Melo, V. V. (2014). Kaizen Programming. In *Proceedings of the 2014 Conference on Genetic and Evolutionary Computation* (pp. 895–902). New York, NY, USA: ACM. DOI: https://doi.org/10.1145/2576768.2598264

Deb, S. (2017). Analytical ideas to improve daily demand forecasts: A case study. *Lecture Notes in Computer Science*, 10192, 23–32. DOI: https://doi.org/10.1007/978-3-319-54430-4_3

Dewhurst, M., & Willmott, P. (2014). Manager and machine: The new leadership equation. *McKinsey Quarterly*, 4, 1–8.

Doward, J. (2017). *Organic food sales soar as shoppers put quality before price* | Environment | The Guardian. Retrieved May 15, 2018, disponível em: https://www.theguardian.com/environment/2017/feb/19/sales-of-organic-food-soar-fruit-vegetables-supermarkets

Draganska, M., Klapper, D., & Villas-Boas, S. B. (2010). A larger slice or a larger pie? An empirical investigation of bargaining power in the distribution channel. *Marketing Science*, 29(1), 57–74. DOI: https://doi.org/10.1287/mksc.1080.0472

Dunkley, B., Helling, A., & Sawicki, D. S. (2004). Accessibility versus scale: examining the tradeoffs in grocery stores. *Journal of Planning Education and Research*, 23(4), 387–401. DOI: https://doi.org/10.1177/0739456X04264890

Estabrooks, A., Jo, T., & Japkowicz, N. (2004). A multiple resampling method for learning from imbalanced data sets. *Computational Intelligence*, 20(1), 18–36. DOI: https://doi.org/10.1111/j.0824-7935.2004.t01-1-00228.x

Fadahunsi, A. (2012). The growth of small businesses: Towards a research agenda. *American Journal of Economics and Business Administration*, 4(1), 105. DOI: https://doi.org/10.3844/ajebasp.2012.105.115

Farhadi, H. (2018). *Machine Learning: Advanced Techniques and Emerging Applications*. BoD--Books on Demand.

Gil, J. M., Gracia, A., & Sanchez, M. (2000). Market segmentation and willingness to pay for organic products in Spain. *The International Food and Agribusiness Management Review*, 3(2), 207–226. DOI: https://doi.org/10.1016/S1096-7508(01)00040-4

Gilbert, R. J. (2015). E-books: A tale of digital disruption. *Journal of Economic Perspectives*, 29(3), 165–184. DOI: https://doi.org/10.1257/jep.29.3.165

Gordon, W. L., & Key, J. R. (1987). Artificial intelligence in support of small business information needs. *Journal of Systems Management*, 38(1), 24.

Ha, A. Y. (1997). Inventory rationing in a make-to-stock production system with several demand classes and lost sales. *Management Science*, 43(8), 1093–1103.

Isidore, C., Wattles, J., & Kavilanz, P. (2018). *Toys "R" Us will close or sell all US stores*. Retrieved May 15, 2018, disponível em: http://money.cnn.com/2018/03/14/news/companies/toys-r-us-closing-stores/index.html

Jones, C. (2004). An alternative view of small firm adaptation. *Journal of Small Business and Enterprise Development*, 11(3), 362-370. DOI: https://doi.org/10.1108/14626000410551618

Kolassa, S. (2013). Forecasting and optimisation for big data: Lessons from the retail business. In OR55 Keynotes and Extended Abstracts - *55th Conference of the Operational Research Society*, 33–35, disponível em: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84887836974&partnerID=40&md5=4a2ff7c1644d89c6f3d3df3073c95d5f

Lee, H. L., & Whang, S. (2000). Information sharing in a supply chain. *International Journal of Manufacturing Technology and Management*, 1(1), 79–93. DOI: https://doi.org/10.1016/j.proeng.2012.06.258

Lensink, R., Van Steen, P., & Sterken, E. (2005). Uncertainty and Growth of the Firm. *Small Business Economics,* 24(4), 381–391. DOI: https://doi.org/10.1007/s11187-005-7121-z

Leswing, K. (2017). *Amazon Is Buying Whole Foods-Here's Amazon's Vision for the Grocery Store of the Future*. Business Insider, disponível em: https://www.businessinsider.com/amazon-go-grocery-store-future-photos-video-2017-6

Li, Y., Su, Z., Liu, Y., & Li, M. (2011). Fast adaptation, strategic flexibility and entrepreneurial roles. Chinese *Management Studies*, 5(3), 256–271. DOI: https://doi.org/10.1108/17506141111163354

Loebbecke, C., & Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *Journal of Strategic Information Systems*, 24(3), 149–157. DOI: https://doi.org/10.1016/j.jsis.2015.08.002

Love, R. R., & Hoey, J. M. (1990). Management science improves fast-food operations. *Interfaces*, 20(2), 21–29.

McFarlane, F. W. (1984). *Information technology changes the way you compete*. Harvard Business Review, Reprint Service.

Moore, C. W. (2008). *Managing small business: An entrepreneurial emphasis*. Cengage Learning EMEA.

Müller, V. C., & Bostrom, N. (2016). Future progress in artificial intelligence: A survey of expert opinion. In *Fundamental issues of artificial intelligence* (pp. 555–572). Springer.

Ramentol, E., Verbiest, N., Bello, R., Caballero, Y., Cornelis, C., & Herrera, F. (2012). SMOTE-FRST: a new resampling method using fuzzy rough set theory. In *Uncertainty Modeling in Knowledge Engineering and Decision Making* (pp. 800–805). World Scientific.

Slimani, I., El Farissi, I., & Achchab, S. (2017). Configuration and implementation of a daily artificial neural network-based forecasting system using real supermarket data. *International Journal of Logistics Systems and Management*, 28(2), 144–163. DOI: https://doi.org/10.1504/IJLSM.2017.086345

Slimani, I., Farissi, I. E., & Al-Qualsadi, S. A. (2016). Configuration of daily demand predicting system based on neural networks. In Proceedings of the *3rd IEEE International Conference on Logistics Operations Management*. DOI: https://doi.org/10.1109/GOL.2016.7731709

Soper, T. (2017). *Amazon reports $1.3B in physical store sales, breaking out brick-and-mortar business for first time, still dwarfed by $26.4B online sales*. GeekWire, disponível em: https://www.geekwire.com/2017/amazon-adds-physical-stores-segment-earnings-report-expands-brick-mortar-footprint/

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., … Horvitz, E. (2016). *Artificial intelligence and life in 2030: One hundred year study on artificial intelligence*. Standford University, disponível em: https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai_100_report_0831fnl.pdf

Taylor, J. W. (2011). Multi-item sales forecasting with total and split exponential smoothing. *Journal of the Operational Research Society*, 62(3), 555–563. DOI: https://doi.org/10.1057/jors.2010.95

Taylor, K., & Hanbury, M. (2018). *Amazon is threatening these 8 industries*. Business Insider. Retrieved May 15, 2018, disponível em: http://www.businessinsider.com/amazon-is-killing-these-7-companies-2017-7#department-stores-3

Thompson, G. D. (1998). Consumer demand for organic foods: what we know and what we need to know. *American Journal of Agricultural Economics*, 80(5), 1113–1118. DOI: https://doi.org/10.2307/1244214

Tu, J. I. (2016). *Costco gets creative to meet shoppers' huge appetite for organics*. The Seattle Times. Retrieved May 15, 2018, disponível em: https://www.seattletimes.com/business/retail/costco-gets-creative-to-meet-shoppers-huge-appetite-for-organics/

Van Doorn, J., & Verhoef, P. C. (2011). Willingness to pay for organic products: Differences between virtue and vice foods. *International Journal of Research in Marketing*, 28(3), 167–180. DOI: https://doi.org/10.1016/j.ijresmar.2011.02.005

Wingfield, N., & de la Merced, M. (2017). *Amazon to buy Whole Foods for 13.4 billion*. The New York Times.

**Como citar este artigo**

Nascimento, A. M.; de Melo, V. V.; Muller Queiroz, A. C.; Brashear-Alejandro, T.; & Meirelles, F. de S. (2020). Inteligência artificial aplicada a pequenas empresas: o uso da engenharia automática de recursos e do aprendizado de máquina para um planejamento mais preciso. *Revista de Contabilidade e Organizações*, 14:e171481. DOI: http://dx.doi.org/10.11606/issn.1982-6486.rco.2020.171481

# Apêndice

$$F1 = \frac{1}{log\left(log\left(\sqrt{DIASEMANA}\right)\right)}$$

$$F2 = \frac{1}{log\left(log\left(\sqrt{MÊS}\right)\right)}$$

$$F3 = \frac{1}{log\left(\frac{DIASEMANA}{log(DAIASEMANA)}\right)}$$

$$F4 = \frac{FERIADO}{log\left(\frac{DIASEMANA}{log(DIASEMANA)}\right)}$$

$$F5 = \frac{1}{2}\left[DIASEMANA + \left(DIASEMANA - FERIADO * \left(\frac{DIAMÊS + \frac{E(VENTOCLIMA + AbsDIAMÊS * DIASEMANA * UMIDADE)}{2}}{2}\right)\right)\right]$$

$$F6 = DIASEMANA - log\left(\frac{1}{2} * \left(MÊS + \left(\sqrt{\frac{1}{FERIADO + \sqrt{log(abs(TEMPMAX))}}}\right)\right)\right)$$

$$F7 = abs\left(\frac{\left(\frac{log(DIASEMANA)}{DIAMÊS} - absTEMPMIN - FERIADO\right) * UMIDADE * MÊS}{DIAMÊS}\right)$$

$$F8 = \frac{1}{DIASEMANA * \frac{TEMPMIN + EVENTOCLIMA}{2}}$$

$$F9 = \frac{1}{log\left(\sqrt{\frac{DIAMÊS}{\frac{\left(TEMPMIN + \left(\frac{TEMPMIN}{TEMPMAX^2}\right)\right)}{2}}}\right)}$$

$$F10 = \frac{DIAMÊS}{\left(DIASEMANA + abs\left(\frac{1}{abs\left(\frac{MÊS}{TEMPMIN}\right)}\right)\right)}$$